STORET Analytical Tools – Concepts and Cross Program Linkages.

2014 National Monitoring Conference
April 28 – May 2, 2014

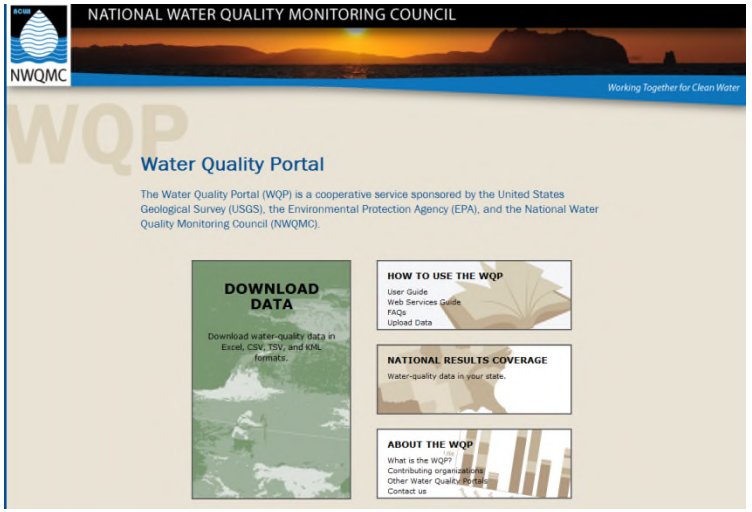
Presenter: **Amy Wesley-Snider (RTI)**

Co-authors: **Danette Boezio, Jay Rineer, Breda Munoz, Peter Ilieve, Michele Eddy, Phillip Jones (RTI)**

3040 Cornwallis Road ■ Research Triangle Park, North Carolina, USA 27709

Presentation Outline

- Data Discovery
  - STORET Repository
  - National Water Quality (NWQ) Portal
  - Additional Data Discovery Functions
- Automated Assessment Tools
  - STORET Trends Analysis Tool
    - Seasonal Kendall Trend
  - Clean Water Act Reporting
  - National Level Screening
  - Exploratory Data Analysis (EDA)
  - National Aquatic Resources Surveys (NARS)
    - Statistical Analysis Tools
  - Additional Functions
    - Upstream/Downstream Summarization
    - NHDPlus Catchments
    - Eco-flows Calculation



Data Discovery

STORET and The National Water Quality (NWQ) Portal

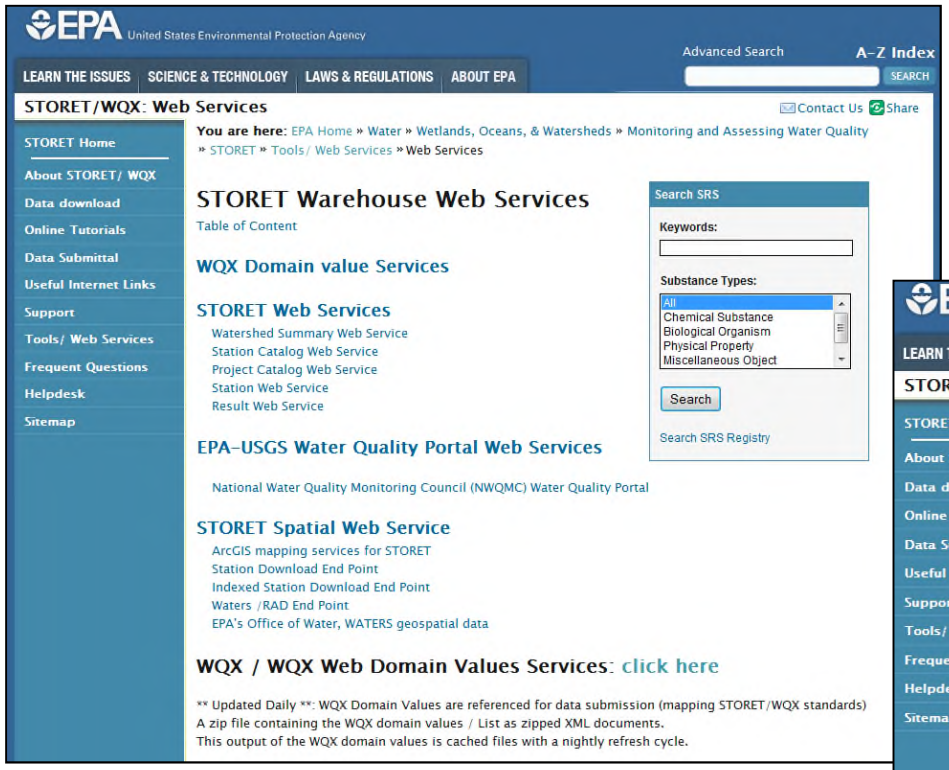
STORET

- STOrage and RETrieval Data Warehouse http://www.epa.gov/storet/
- National repository for water quality monitoring data collected by water resource management groups.

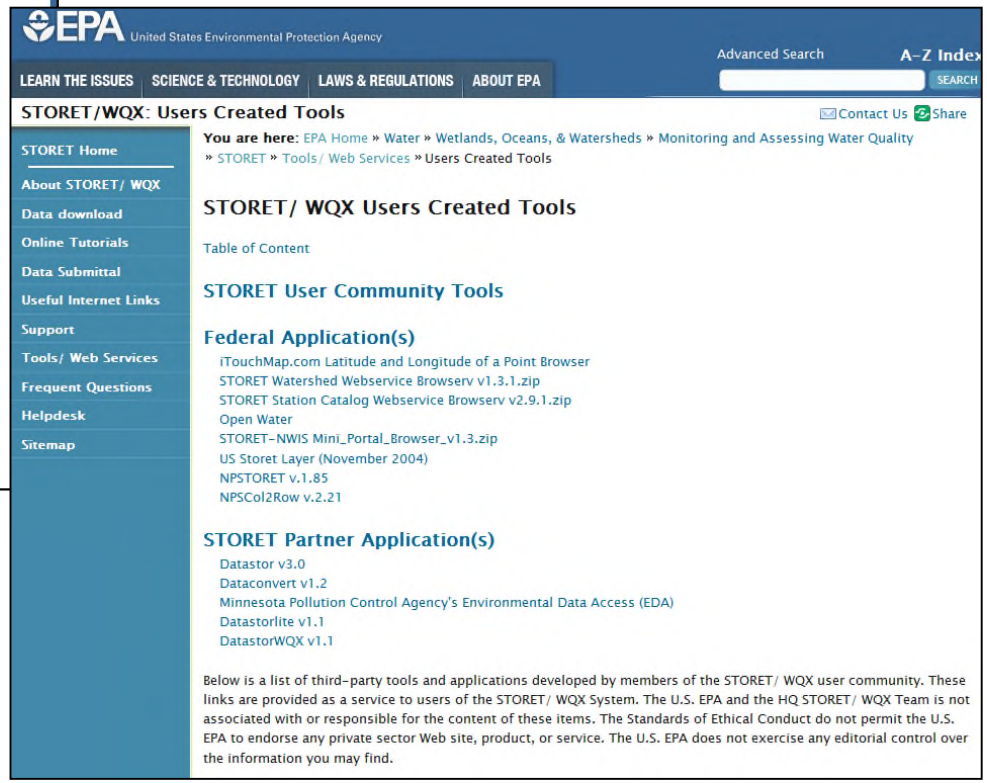
NWQ Portal

- Cooperative service sponsored by USGS, EPA and NWQMC
- Provides access to data stored in large Water quality Databases
  - USGS NWIS
  - STORET

Additional Data Discovery Functions



http://www.epa.gov/storet/tools.html



Automated Assessment Tools

Automated Assessment Tools

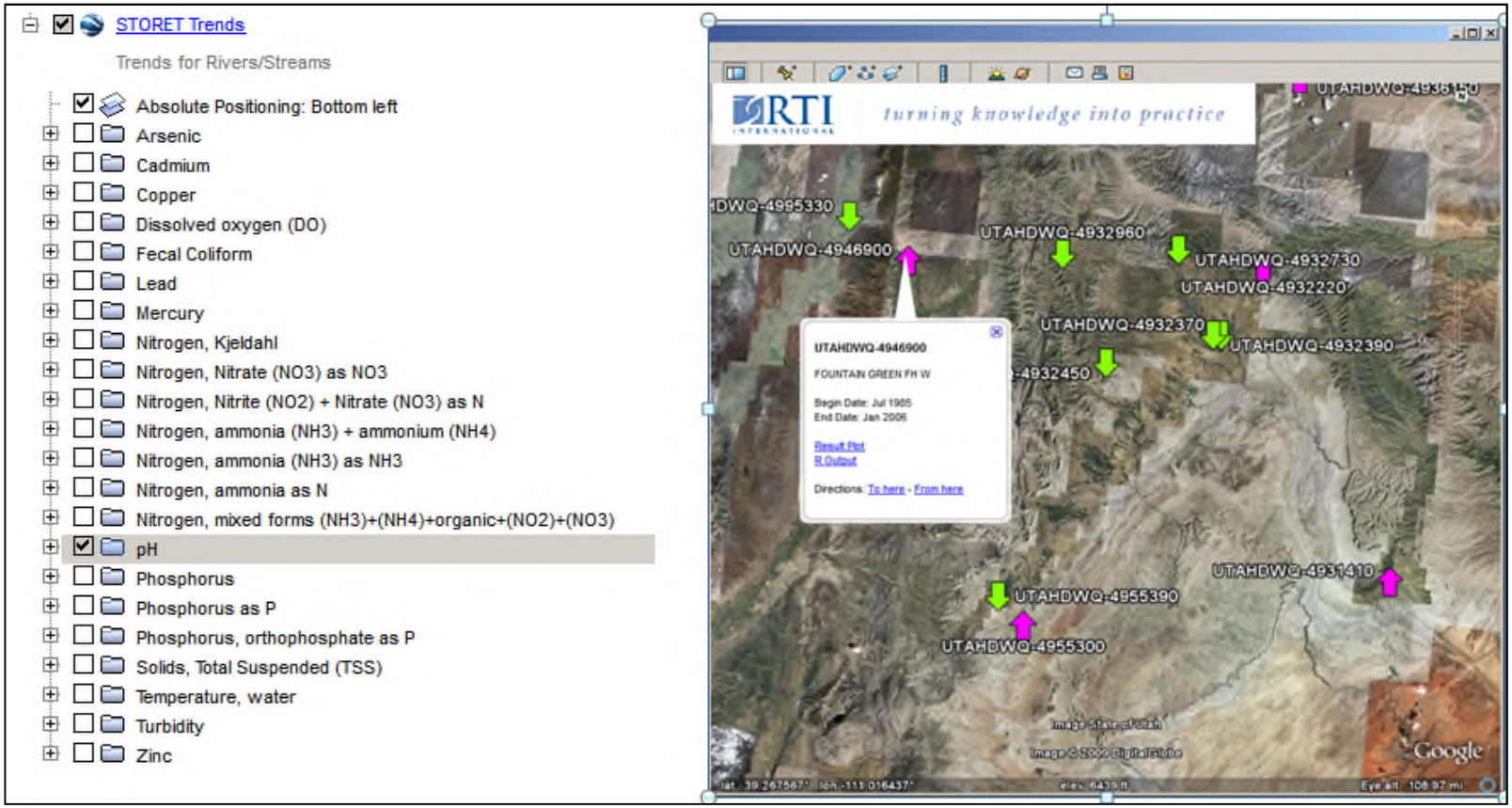
1. STORET Trends Analysis
2. Clean Water Act (CWA) Reporting
3. National Level Screening
4. Exploratory Data Analysis (EDA)
5. National Aquatic Resources Surveys (NARS) Statistical Analysis

1. STORET Trends Analysis Tool

- Goal: Provide users with the ability to statistically identify trends in STORET Data.
- Pilot was run on a static copy of the STORET database.
  - Conducted a Seasonal Kendall Trend analysis for 22 characteristic types at all STORET stations.
    - Utilized open source “R” for this analysis.

1. STORET Trends Analysis Tool cont.

STORET Characteristics used for the Pilot Study

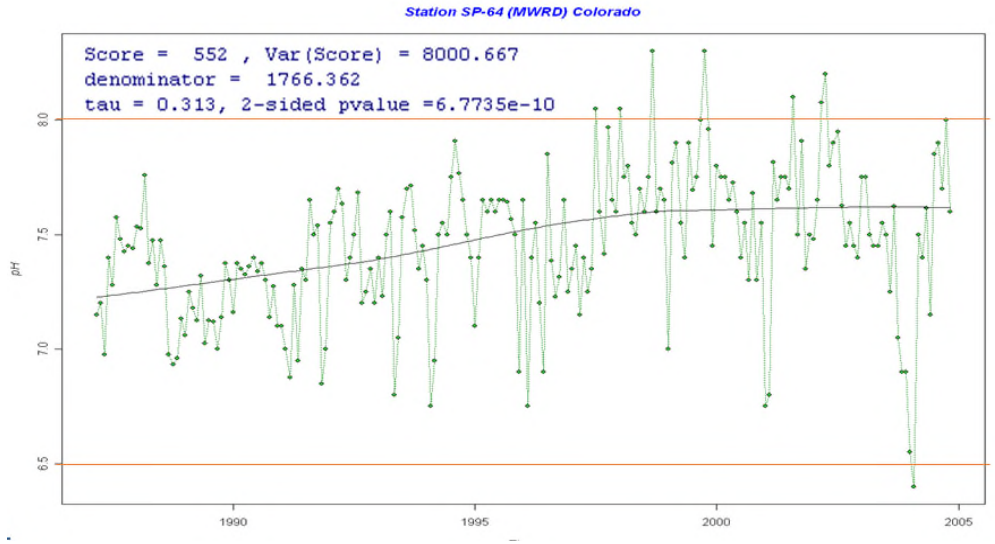


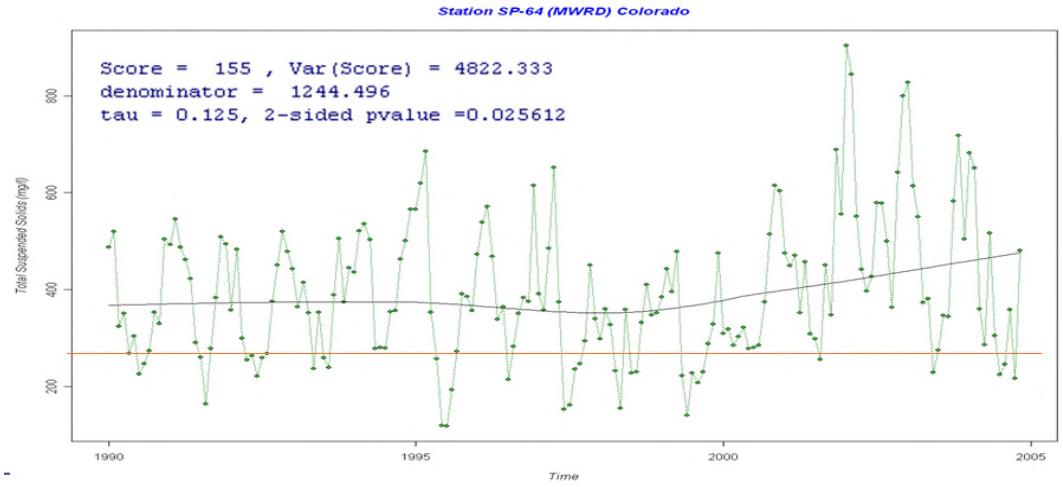
1. STORET Trends Analysis Tool cont.

- Stations that did not have an adequate period of records to support a trends analysis were removed.
  - Created a script to filter for stations having a time series with at least four measurements collected periodically for at least five years.
- A minimum of 20 measurements was determined to provide for the investigation of serial correlation (Yue and Wang, 2004)
- Produced trend plots and scores for each parameter.

1. STORET Trends Analysis Tool cont.

Trend Plot for pH produced by the STORET Trends Analysis Tool

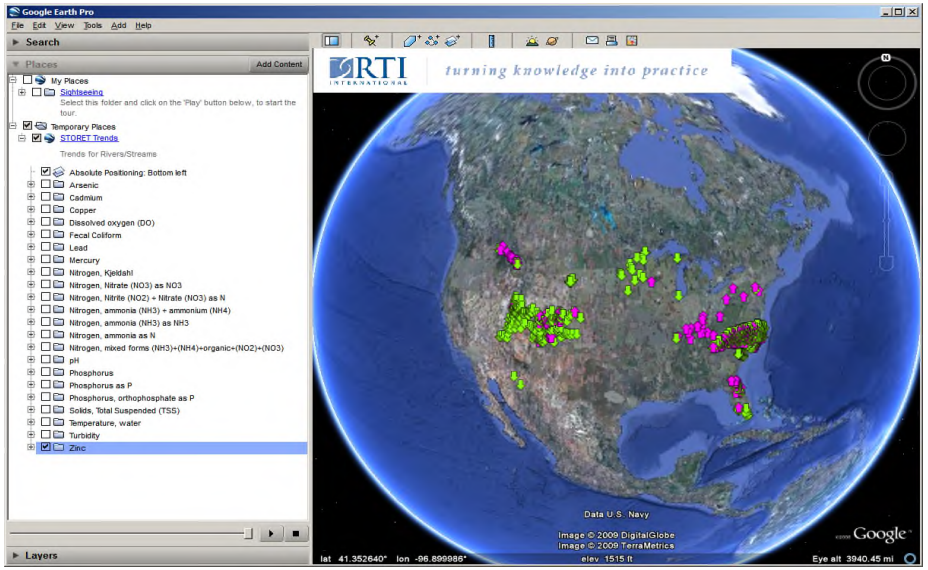




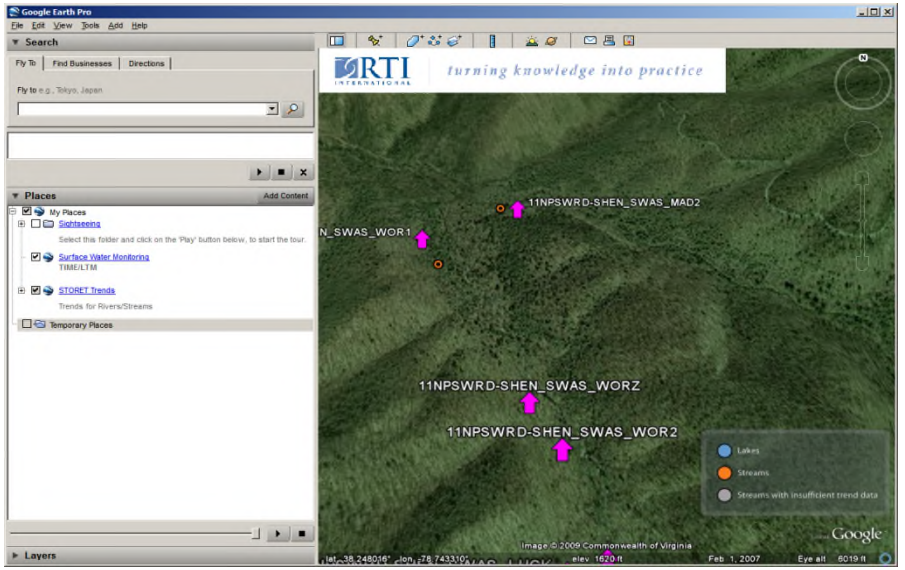
Trend Plot for TSS produced by the STORET Trends Analysis Tool

1. STORET Trends Analysis Tool cont.

- Tool produced a Google KML map interface for examining national characteristic trends
- Users can use interface to visualize the trend nationally or drill in on a specific station of interest.



STORET Trends Analysis Tool, National View



STORET Trends Analysis Tool, Drill Down View

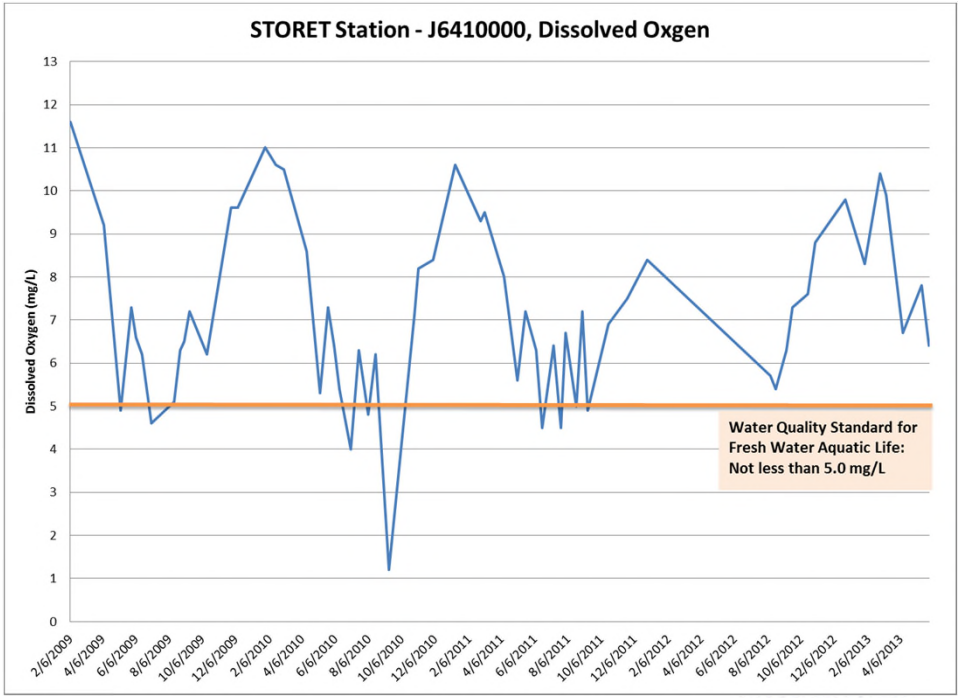
2. CWA Reporting Tools

- Compare STORET/NWQ Portal data to a value (e.g., a WQS) for a parameter of interest.
- Tailored to state-specific needs: waterbody-specific water quality criteria and standards.
- Screening exceedances could be flagged, and samples with results of concern could be highlighted to steer a user to particular areas that may have water quality problems.

2. CWA Reporting Tools Cont.

Little Creek (West Side): NC27-86-2-4

Total # of Results	Total # of Exceedences	Min	Max	Average
60	8	1.2	11.6	7.185



3. National Level Screening

- Under CWA 304(a) EPA has established National Criteria to provide guidance for states/tribes to use in adopting Water Quality Standards.
- Create a tool to assist in screening against the national recommended water quality criteria.
- Example: Identify all observations of nitrogen/ phosphorus that exceed the ecoregional nutrient criteria.

3. National Level Screening Cont.

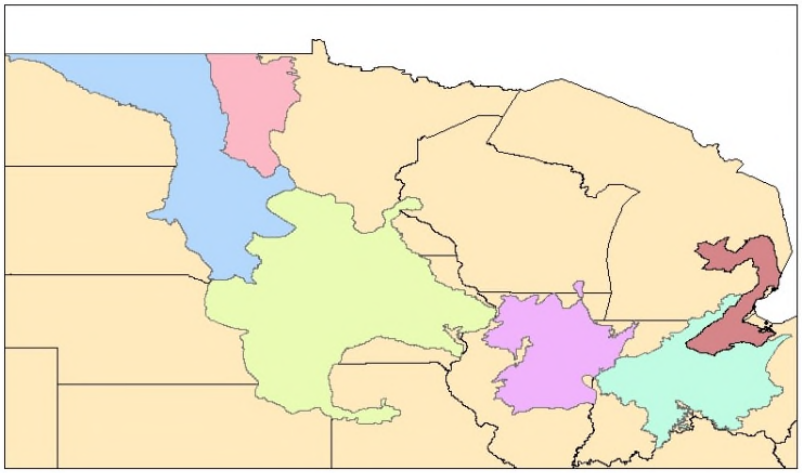
- EPA Developed Ecoregional Criteria for Total Phosphorus, Total Nitrogen, Chlorophyll *a* and Water Clarity.
- For this example Ecoregion VI, #55 *Eastern Corn Belt Plains* was used.

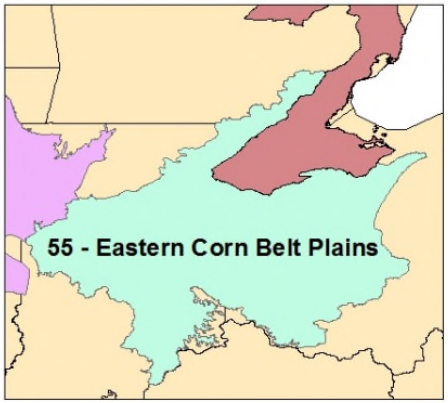
Table 3e. Reference conditions for level III ecoregion 55.

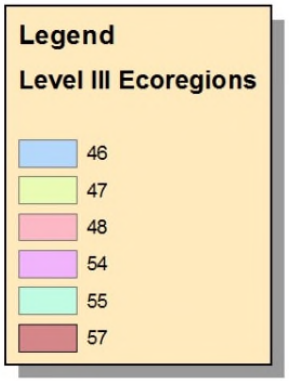
Parameter	No. of Streams N ++	Reported values		25th Percentiles based on all seasons data for the Decade	Reference Streams **
		Min	Max	P25-all seasons*	P75 - all seasons
TKN (mg/L)	198	0.05	3.5	0.4	
$NO_2 + NO_3$ (mg/L)	219	0.025	8.13	1.60	
TN (mg/L) - calculated	NA	0.075	11.63	2	
TN (mg/L) - reported	2 z	3.63	3.78	3.63	
TP (ug/L)	225	10	1820	62.5	
Turbidity (NTU)	1 z	10.4	10.4	10.4 zz	
Turbidity (FTU)	12	3.3	50.65	9.21	
Turbidity (JCU)	1 z	28	28	28 zz	
Chlorophyll *a* (ug/L) -F	—	—	—	—	
Chlorophyll *a* (ug/L) -S	8	4.32	19.24	6.62	
Chlorophyll *a* (ug/L) -T	8	6.67	22.72	7.99	
Periphyton Chl *a* (mg/m^2)	—	—	—	—	—

3. National Level Screening Cont.

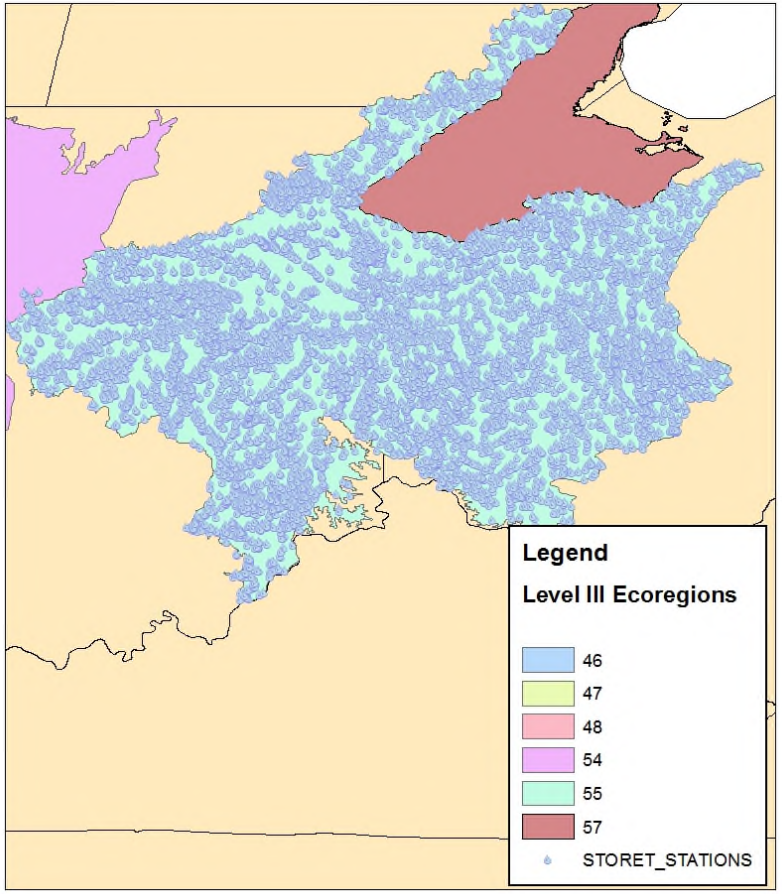
Aggregate Nutrient Ecoregion 6







Level III Ecoregion - 55: Eastern Corn Belt Plains
River and Stream Monitoring Stations



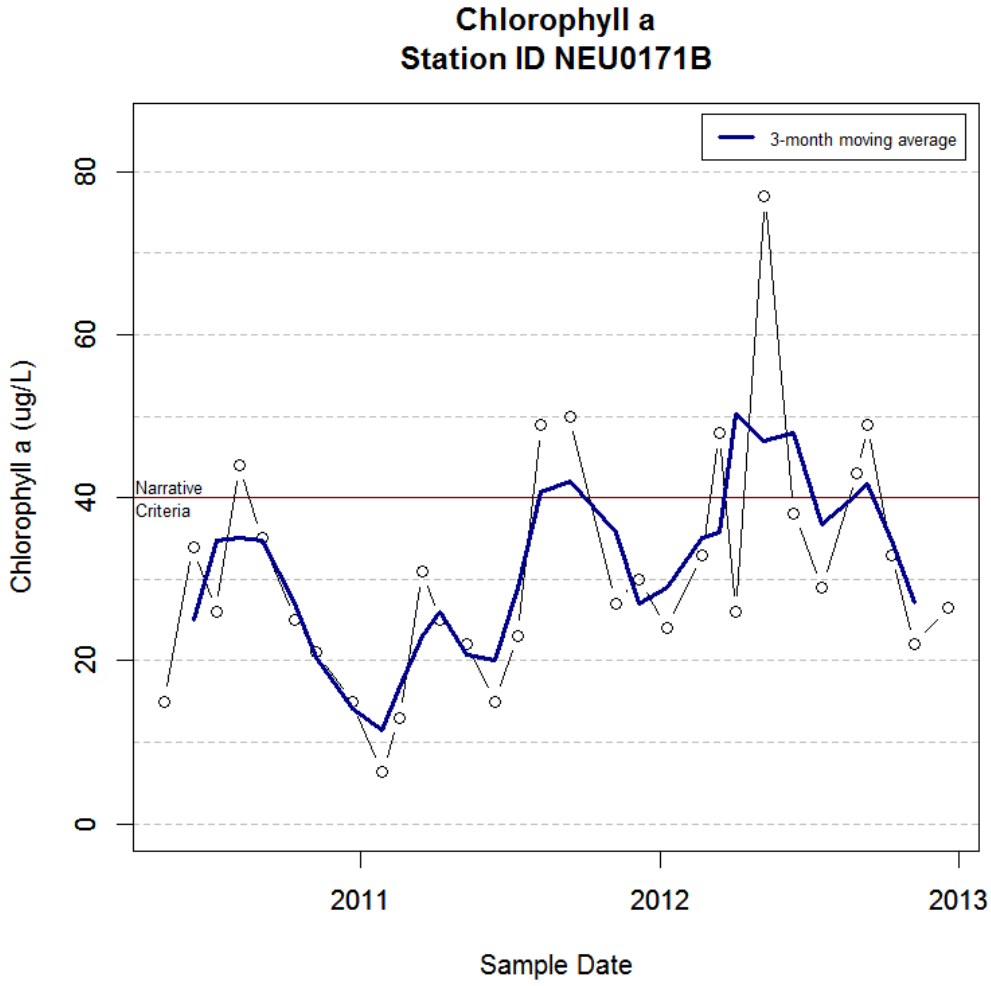
3. National Level Screening Cont.

- For this ecoregion there are over 12,000 stations
- Screening Tool would allow user to:
  - Define parameter,
  - Define date range,
  - Select statistics for data (e.g. number of samples that exceed the reference conditions for parameter)
- Screening tool would return results in user defined format (.xls, .csv, etc..)

4. Exploratory Data Analysis

- Exploratory data analysis (EDA)
  - Help identify degrading waters that may need additional protection.
  - Highlight improving waters showing progress towards restoration.
- EDA tools can be used to:
  - Examine parameter trends over time.
  - Identify relationships between parameters and other variables.
  - Estimate the probability of threshold exceedances.
  - Characterize central tendency (mean, median) and variability of parameter values.
  - Compare outputs described above by: station, watershed, county, eco-region, sample month, season, etc.
- Tool outputs can be graphical and/or tabular

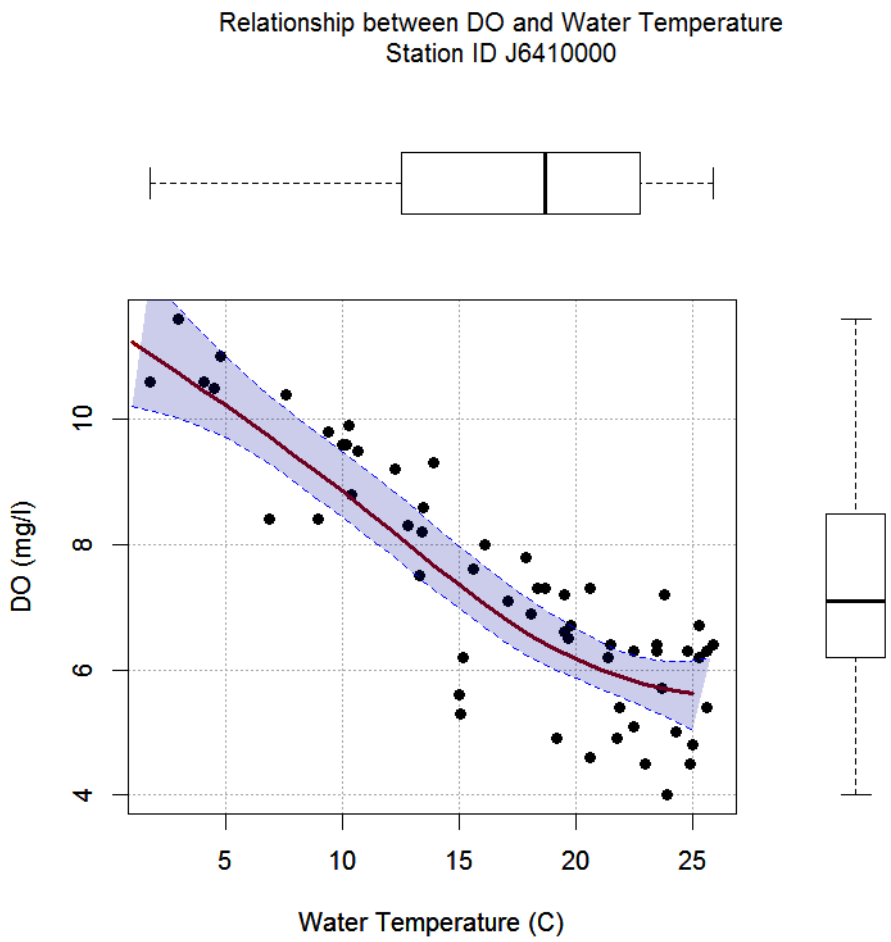
4. Exploratory Data Analysis-Time Series Data



Time series data:

- Identify potential trends over time relative to WQ criteria or other thresholds.
- Add moving average to help clarify trends obscured by noisy (high variability) data.

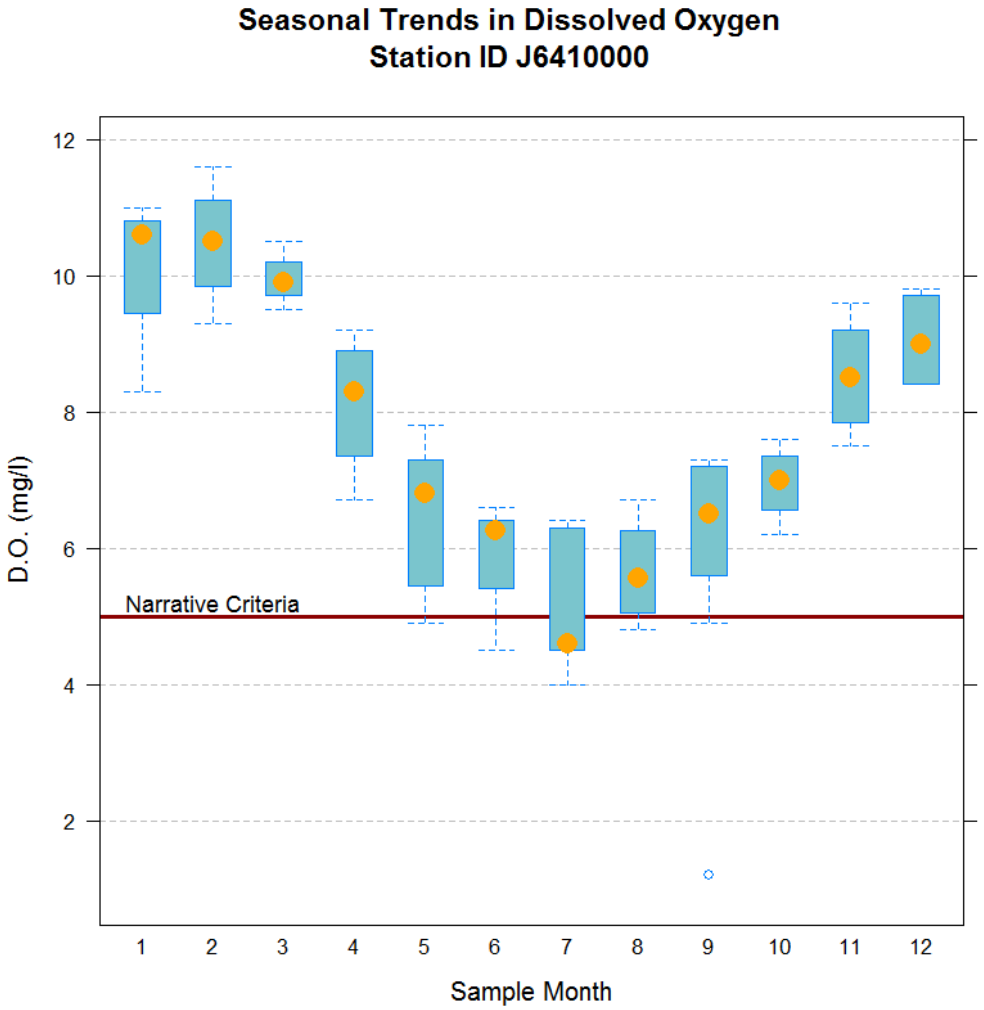
4. Exploratory Data Analysis- Scatterplots/Trend lines



Scatterplots with trend lines:

- Identify the direction, shape, and strength of relationships between variables
- Estimate and display prediction intervals.

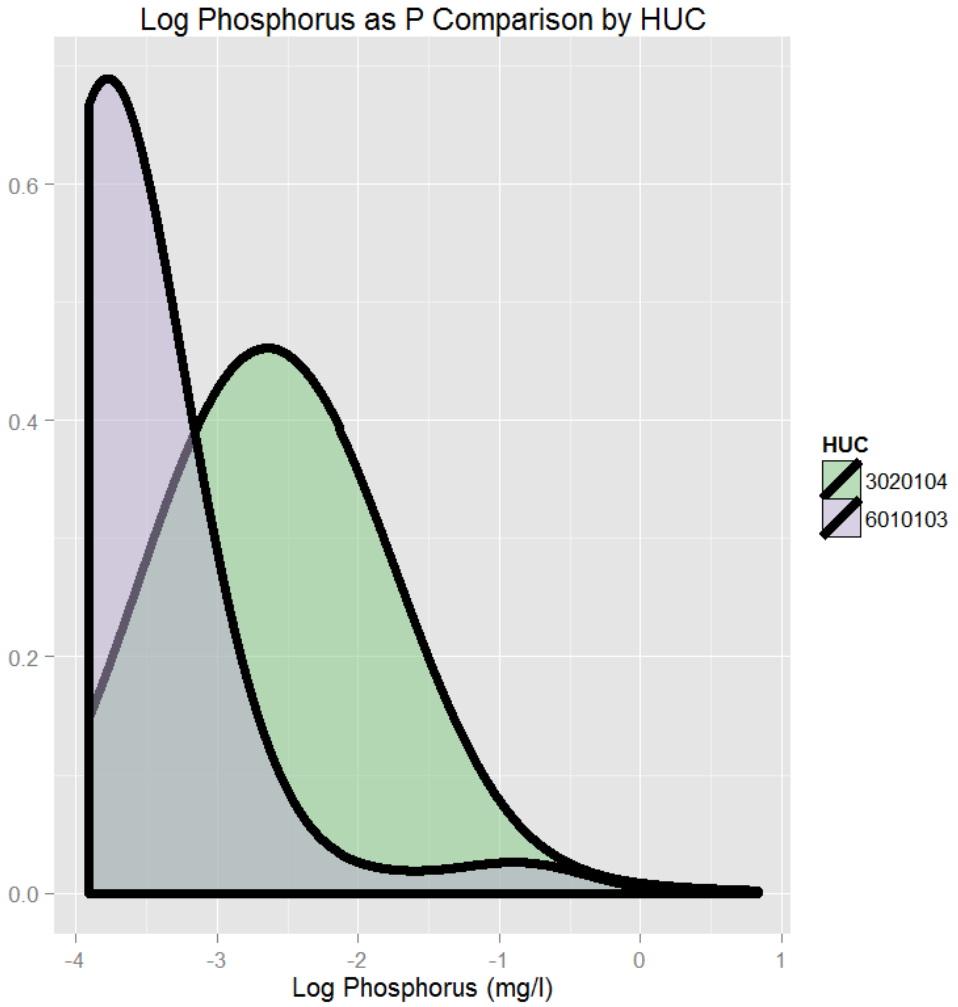
4. Exploratory Data Analysis - Boxplots



Boxplots:

- Show median values, inner quartile range (25th to 75th percentile), outer quartiles and outliers.
- Identify potential spatial or temporal trends.

4. Exploratory Data Analysis– Density Estimates



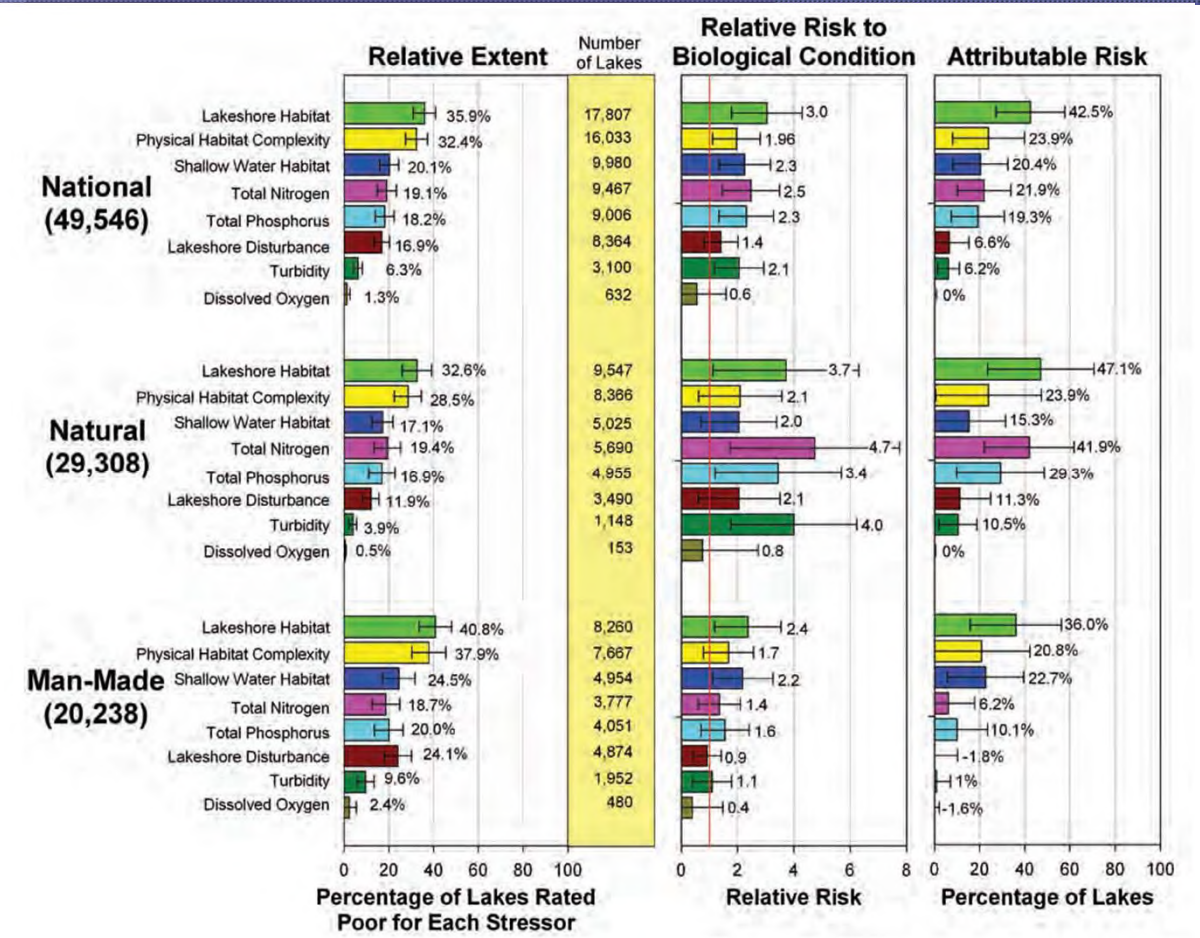
Density plots:

- Display central tendency (mean or median) and spread (variability) of parameter.
- Compare different spatial or temporal units (i.e., is one HUC more likely to experience an exceedance?).
- Identify skewed distributions (i.e., higher probability of extreme values).

5. NARS Tool

- Develop a tool to calculate relative and Attributable Risk for the NARS data set.
  - Relative risk: examines the relationship between a stressor and the response in a measure of biological condition.
  - Attributable risk: builds on relative risk by looking at the likelihood that poor biological condition will exist when stressor levels are rated poor, and at the extent to which those poor conditions exist in the assessed region.

5. NARS Tool



Relative extent, relative risk and attributable risk has been used to assess the condition of the nation's lakes (US EPA, 2009)

Additional Functions

- **Upstream/downstream summarization**: Select a pour point and summarize data upstream. Or once a station is selected, select any other stations with same parameter downstream for X km. Results could be plotted on the same graph(s) for comparison.
- **NHDPlus Catchments**: Report out the NHDPlus catchments associated with monitoring stations.
- **Ecoflows Calculation**: STORET contains biological and habitat data that could be useful for site-specific or cross-sectional ecological flow analysis. (Note: RTI has demonstrated the potential for such cross-sectional analyses using data from North Carolina)

Conclusions

- A range of nationwide analytical could be developed to access the wealth of data within the STORET Data Warehouse.
- Streamline the process of compiling relevant information for state users and EPA decision makers.
- Reinforce interest in the STORET Data Warehouse and bolster the user community's interest in providing data.
- Utilize Open Source software for development:
  - Eliminates the costs associated with licensing fees and/or costly updates due to software version changes.
  - Allows for easier code sharing for customization by jurisdictions.

For More Information

Amy Wesley-Snider (RTI International)

919-316-3711, **alwesley@rti.org**